

# Kiến trúc kho dữ liệu mở ngành Tài nguyên và Môi trường tích hợp lên Hệ tri thức Việt số hóa

○ ĐẶNG THỊ THU TRANG

PHẠM MINH TRƯỜNG, NHÂM NGỌC TÂN

Cục Công nghệ thông tin và Dữ liệu Tài nguyên và Môi trường

*Ngành Tài nguyên và Môi trường là ngành điều tra cơ bản, kết quả chủ yếu là dữ liệu với một số đặc thù riêng như dữ liệu quan trắc, dữ liệu không gian, dữ liệu viễn thám,... những dữ liệu này có vai trò rất quan trọng trong công cuộc quản lý, xây dựng và phát triển đất nước. Việc mở dữ liệu mang lại rất nhiều lợi ích, tạo động lực cho sự phát triển kinh tế và khoa học thông qua sự thuận tiện hơn khi tiếp cận các thông tin hợp nhất từ nhiều kho dữ liệu khác nhau của các lĩnh vực trong ngành Tài nguyên và Môi trường. Xây dựng kho dữ liệu mở Ngành đáp ứng được nhu cầu khai thác thông tin dữ liệu ngành Tài nguyên và Môi trường, phù hợp với định hướng phát triển của Chính phủ trong bối cảnh Cách mạng Công nghiệp lần thứ 4.*

## Mô hình tổng thể Kho dữ liệu mở ngành Tài nguyên và Môi trường

Kho dữ liệu mở ngành TN&MT là một nơi quản lý, lưu trữ và khai thác thông tin về dữ liệu mở trong ngành TN&MT. Nó bao gồm các thành phần dữ liệu chính sau:

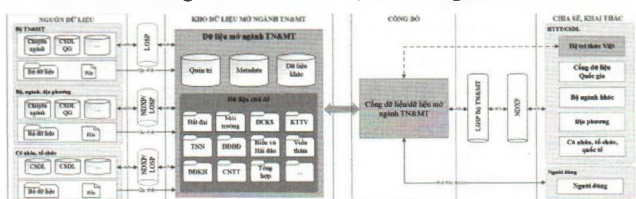
**Dữ liệu chủ đề:** Là các bộ dữ liệu, các tệp dữ liệu được phân chia theo từng chủ đề của từng lĩnh vực trong ngành TN&MT như: Đất đai, Môi trường, ĐC&KS, KTTV,... Bên cạnh đó, còn có các chủ đề về CNTT, Tổng hợp,...

**Dữ liệu quản trị:** Là thông tin dữ liệu để quản lý và vận hành kho dữ liệu mở ngành TN&MT như: Dữ liệu về phân quyền người dùng, dữ liệu về quá trình cung cấp dữ liệu, dữ liệu về cấu hình hệ thống,...

**Dữ liệu metadata:** Là các siêu dữ liệu mô tả thông tin các bộ dữ liệu có trong kho dữ liệu mở ngành TN&MT. Các siêu dữ liệu này được quy định theo các tiêu chuẩn quốc tế đối với từng loại dữ liệu như: ISO19115, ISO19139, DublinCore, Marc21, DCAT,...

**Dữ liệu khác:** Là các loại dữ liệu khác những loại dữ liệu đã liệt kê ở trên, bao gồm: Dữ liệu từ các tổ chức, cá nhân đóng góp, Dữ liệu danh mục dùng chung,...

Mô hình tổng thể kho dữ liệu mở ngành TN&MT



Dữ liệu mở ngành TN&MT được thu thập từ nhiều nguồn dữ liệu khác nhau thông qua việc kết nối, tích hợp từ các HTTT/CSDL hoặc được cập nhập trực tiếp. Các nguồn dữ liệu chính gồm: Các đơn vị trong Bộ TN&MT; các bộ, ngành, địa phương; các cá nhân, tổ chức, quốc tế.

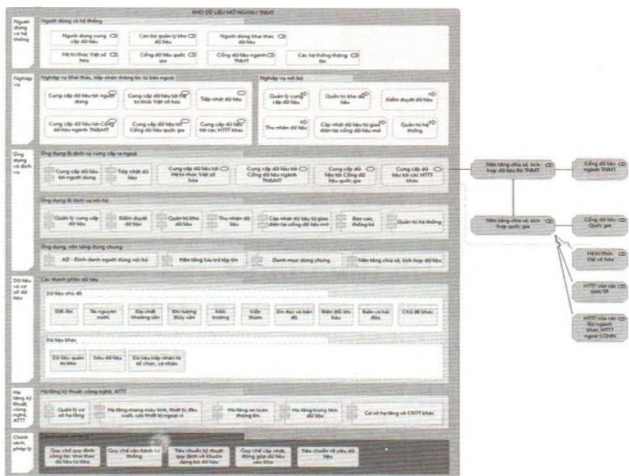
Dữ liệu mở ngành TN&MT sẽ được công bố, cung cấp miễn phí thông qua Cổng dữ liệu/dữ liệu mở ngành TN&MT. Tại đây, người dùng có thể tra cứu, khai thác thông tin dữ liệu mở ngành TN&MT thông qua các giao diện chức năng của Cổng dữ liệu/dữ liệu mở ngành TN&MT. Ngoài ra, dữ liệu mở ngành TN&MT cũng được chia sẻ cho các HTTT/CSDL khác có nhu cầu bằng việc xuất bản các dịch vụ dữ liệu (API) tích hợp vào Nền tảng chia sẻ, tích hợp dữ liệu của Bộ TN&MT (LGSP) và Nền tảng trao đổi dữ liệu quốc gia (NDXP).

## Kiến trúc Kho dữ liệu mở ngành Tài nguyên và Môi trường

Kiến trúc Kho dữ liệu mở ngành TN&MT được xây dựng bằng việc chọn lọc, kết hợp nhiều phương pháp xây dựng kiến trúc để đáp ứng nhu cầu của Bộ TN&MT, nhu cầu khai thác, sử dụng thông tin, dữ liệu TN&MT của bộ, ngành, địa phương, nhu cầu thông tin, dữ liệu TN&MT công bố trên Hệ tri thức Việt số hóa,...

Mô tả các thành phần trong mô hình kiến trúc kho dữ liệu mở ngành TN&MT như sau:

**Người dùng hệ thống:** Bao gồm tất cả người dùng, hệ thống tham gia trong quá trình xây dựng, vận hành, cung cấp, khai thác dữ liệu từ Kho dữ liệu mở ngành TN&MT. Phân loại và mô tả chi tiết được cụ thể tại bảng mô tả người dùng và vai trò.



**Tầng nghiệp vụ:** Thể hiện các quy trình, nghiệp vụ được tin học hóa trong Kiến trúc Kho dữ liệu mở ngành TN&MT, nhằm thể hiện rõ hơn vai trò người dùng đối với hệ thống gắn với các nghiệp vụ cụ thể, đảm bảo tính logic trong kiến trúc.

**Tầng ứng dụng và dịch vụ:** Bao gồm các nhóm chức năng và dịch vụ về cập nhật, quản lý, kiểm duyệt, cung cấp, chia sẻ dữ liệu mở,... Ứng dụng và dịch vụ được xây dựng trên cơ sở kế thừa, dùng chung các nền tảng dùng chung sẵn có được quy định tại Kiến trúc CPĐT ngành TN&MT.

**Tầng dữ liệu và cơ sở dữ liệu thể hiện các thành phần dữ liệu có trong Kho dữ liệu mở ngành TN&MT.** Các nhóm dữ liệu bao gồm: (1) Nhóm dữ liệu chủ đề: Là các bộ dữ liệu, các tập dữ liệu được phân chia theo từng chủ đề như Đất đai, Môi trường, ĐC&KS, KTTV, ĐĐ&BD và thông tin địa lý, TNMT Biển và hải đảo, BDKH, TNN, Viễn thám, Tổng hợp,... (2) Dữ liệu quản trị; (3) Dữ liệu metadata; (4) Dữ liệu khác.

Tầng hạ tầng kỹ thuật, công nghệ, ATTT quy định về các giải pháp công nghệ, hạ tầng CNTT, các giải pháp đảm bảo an toàn thông tin phục vụ xây dựng, triển khai và vận hành Kho dữ liệu mở ngành TN&MT. Kho dữ liệu mở ngành TN&MT được triển khai, vận hành tại Trung tâm Dữ liệu dùng chung của Bộ TN&MT, sử dụng hạ tầng CNTT, các giải pháp an toàn, an ninh sẵn có.

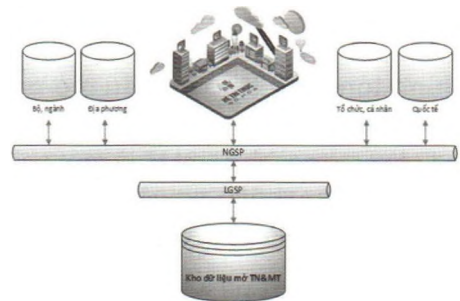
Tầng chính sách và pháp lý xác định khung pháp lý nhằm đảm bảo việc quản lý, duy trì, vận hành, cập nhật, khai thác Kho dữ liệu mở ngành TN&MT.

**Giải pháp tích hợp dữ liệu ngành Tài nguyên và Môi trường lên Hệ tri thức Việt số hóa**

Đề án “Hệ tri thức Việt số hoá” được Thủ tướng Chính phủ phê duyệt theo Quyết định số 677/QĐ-TTg ngày 18/5/2017. Kết quả của đề án được triển khai tại địa chỉ <https://itrithuc.vn>, Hệ tri thức Việt số hóa hứa hẹn sẽ tạo ra một hệ sinh thái toàn diện để tất cả mọi người, nhất là thế hệ trẻ Việt Nam sáng tạo, phát triển các công nghệ tiên tiến trên nền tảng

của dữ liệu lớn, IoT, trí thông minh nhân tạo..., đặc biệt trong thời kỳ Cách mạng Công nghiệp 4.0 đang tác động mạnh mẽ tới mọi lĩnh vực từ khoa học, giáo dục, giao thông, tài chính, sản xuất, y tế,... Đây chính là nền tảng kiến tạo những cơ hội lớn, thực tiễn cho phong trào khởi nghiệp đổi mới sáng tạo của Việt Nam. Đến tháng 12/2019, Hệ tri thức Việt số Hóa đã cung cấp 10.426 Bộ dữ liệu về các lĩnh vực khác nhau, 3,5 triệu câu hỏi và trả lời; 3.177 bài viết trên “cây tri thức”; 7 ứng dụng,... Bên cạnh đó, 12 bộ, ngành đã xác định các dữ liệu có thể đưa lên Hệ tri thức, trong đó có Bộ TN&MT.

Theo Kiến trúc Chính phủ điện tử ngành TN&MT, việc kết nối, liên thông giữa các hệ thống thông tin/cơ sở dữ liệu trong Bộ, giữa các hệ thống thông tin/cơ sở dữ liệu trong Bộ và ngoài Bộ được thực hiện thông qua nền tảng chia sẻ tích hợp quốc gia (NGSP), cấp bộ, ngành, địa phương (LGSP). Việc kết nối, chia sẻ của kho dữ liệu mở ngành TN&MT với Hệ tri thức Việt số hóa được thể hiện như sau:



**Mô hình kết nối Kho dữ liệu mở ngành TN&MT với Hệ tri thức Việt số hóa**

Kho dữ liệu mở ngành TN&MT có thể kết nối, tích hợp dữ liệu 2 chiều với Hệ tri thức Việt số hóa, các hệ thống của các bộ, ngành, địa phương và các hệ thống của các tổ chức cá nhân, quốc tế.

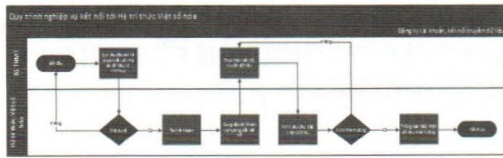
Đối với chiều cung cấp dữ liệu: Kho dữ liệu mở ngành TN&MT sẽ đóng gói dữ liệu thành các bộ dữ liệu (dataset) theo cấu trúc của chuẩn dữ liệu mở, sau đó sẽ xuất bản thành dịch vụ dữ liệu. Dịch vụ dữ liệu này sẽ được tích hợp vào Nền tảng chia sẻ, tích hợp dữ liệu (LGSP) của Bộ TN&MT, sau đó được tích hợp vào Nền tảng chia sẻ, tích hợp dữ liệu quốc gia (NGSP). Các hệ thống có nhu cầu khai thác dữ liệu mở của ngành TN&MT sẽ kết nối, tích hợp vào Nền tảng chia sẻ, tích hợp quốc gia (NGSP) để tích hợp, tra cứu và khai thác thông tin.

Đối với chiều thu nhận dữ liệu: Các hệ thống cung cấp dữ liệu của các Bộ, ngành, địa phương và các hệ thống của các tổ chức cá nhân, quốc tế xuất bản dịch vụ dữ liệu, sau đó tích hợp với Nền tảng chia sẻ, tích hợp dữ liệu quốc gia (NGSP). Nền tảng chia sẻ, tích hợp dữ liệu (LGSP) của Bộ TN&MT sẽ kết nối, tích hợp với Nền tảng chia sẻ, tích hợp dữ liệu quốc gia (NGSP) để cung cấp các dịch vụ dữ



liệu cho Kho dữ liệu mở ngành TN&MT thu nhận dữ liệu vào hệ thống.

Nghiệp vụ đăng ký tài khoản, kết nối và truyền dữ liệu tới Hệ tri thức Việt số hóa thực hiện theo các bước sau:



### Quy trình nghiệp vụ kết nối với Hệ tri thức Việt số hóa

Đầu tiên, Bộ TN&MT cần gửi yêu cầu đăng ký, tạo tài khoản tới Hệ tri thức Việt số hóa bằng công văn. Sau khi tiếp nhận công văn yêu cầu tạo tài khoản kết nối, đơn vị duy trì vận hành Hệ tri thức Việt số hóa có trách nhiệm tiếp nhận và xử lý yêu cầu. Trong trường hợp yêu cầu đăng ký tài khoản được phê duyệt, đơn vị vận hành Hệ tri thức sẽ tạo, cung cấp thông tin tài khoản cùng hướng dẫn kết nối tới Bộ TN&MT. Khi có dữ liệu cần chia sẻ tới Hệ tri thức Việt số hóa, Bộ TN&MT thực hiện kết nối và truyền dữ liệu. Hệ tri thức Việt số hóa sau khi tiếp nhận yêu cầu truyền dữ liệu sẽ xử lý yêu cầu và tiếp nhận dữ liệu. Kết quả xử lý sẽ được gửi lại dưới dạng thông báo tới Bộ TN&MT.

Thông điệp yêu cầu và kết quả trả về phải được chuẩn hóa để hai hệ thống có theo trao đổi thông tin dễ dàng. Do đó, cần các quy định về kết nối chia sẻ dữ liệu tới Hệ tri thức Việt số hóa. Cụ thể như sau:

Để kết nối được với Hệ tri thức Việt số hóa cần có 2 thông tin xác thực là: merchant\_site\_code và secure\_pass. Thông tin này do người quản trị hệ thống Hệ tri thức Việt số hóa cung cấp sau khi hoàn thành quá trình đăng ký.

#### Quy định thông điệp trao đổi:

```

    $api_url = https://dulieu.itrithuc.vn/?com=api&fun=dataset;
    //function POST
    function sendJsonOverPost($data) {
        $content = json_encode($data);
        // open connection
        $ch = curl_init($api_url);
        // set the url, number of POST vars, POST data curl_setopt($ch,
        CURLOPT_URL, $url);
        curl_setopt($ch, CURLOPT_SSL_VERIFYHOST, 0);
        curl_setopt($ch, CURLOPT_SSL_VERIFYPEER, 0);
        curl_setopt($ch, CURLOPT_POST, 1); curl_setopt($ch, CURLOPT_POSTFIELDS,
        $content); curl_setopt($ch, CURLOPT_RETURNTRANSFER, 1);
        curl_setopt($ch, CURLOPT_FOLLOWLOCATION, 1);
        curl_setopt($ch, CURLOPT_HTTPHEADER, array('Content-Type:
        application/json'));
        // execute post
        $result = curl_exec($ch);
        $status = curl_getinfo($ch, CURLINFO_HTTP_CODE);
        $error = curl_error($ch);
        if ($result != '' && $status==200) return $result;
    }
    return false;
  }
  
```

Cấu trúc thông điệp trả về từ Hệ tri thức Việt số hóa:

```

    $data = Array ( [id] => [secure_code] => [secure_pass] => [catid] =>
    [groupid] => [lid]
    => [vid] => [admin_id] => [addtime] => [edittime] => [status] => [title] =>
    [alias] => [author] => [author_email] => [maintainer] => [maintainer_email]
    => [url] => [version] => [notes] => [inhome] => [hitatotal] => [hitscm] =>
    [keywords] => [resource] => Array ( [0] => Array ( [id] => 0 [file] =>
    [link] => [name] => [description]
    => [format] => ) [1] => Array ( [id] => 0 [file] => [link] => [name] =>
    [description] => [format] => ) ) )
  
```

Ví dụ về cấu trúc thông điệp trả về từ Hệ tri thức Việt số hóa:

```

    Array
    (
        [id] => 11111
        [secure_code] => 12345
        [secure_pass] => 7115ee9c7d5d6229d2921a40e899cc11
        [catid] => 17
        [groupid] => 0
        [lid] => 0
        [vid] => 0
        [admin_id] => 22289
        [addtime] => 1524664720
        [edittime] => 1534582102
        [status] => 1
        [title] => del1123
        [alias] => nghiên-cuu-co-che-gial-phong-asen-ra-nuoc-ngam-mo-phong-cac-qua-trinh-khu-yam-khi-tu-nh
        [author] => Trần Hồng Côn, Nguyễn Thị Minh
        [author_email] =>
        [maintainer] => Bộ học Quốc gia Hà Nội
        [maintainer_email] => nhietp@vnu.edu.vn
        [url] => http://repository.vnu.edu.vn/
        [version] =>
        [notes] => Đồng trước sự đe dọa và hiểm họa của tình trạng nhiễm arsen trong nước sinh hoạt và ăn u
        [inhome] => 1
        [keywords] =>
        [resource] => Array
        (
            [0] => Array
            (
                [id] => 0
                [file] => https://csdl.vista.gov.vn/media/publication/306071/Cv85123020028.pdf
                [link] =>
                [name] => 28-Nguyễn Hồng Minh-KYMT 20 năm.pdf
                [description] => Sự yếu hội nghị - Hội thảo (LIC)
                [format] => 1
            )
            [1] => Array
            (
                [id] => 0
                [file] => 2020_08/ky_00357.pdf
                [link] => https://csdl.vista.gov.vn/media/publication/306071/Cv85123020028.pdf
                [name] => KY_00357.pdf
                [description] =>
                [format] => 1
            )
        )
    )
  
```

Xây dựng kiến trúc kho dữ liệu mở ngành TN&MT tích hợp lên Hệ tri thức Việt số hóa đã đáp ứng được mục tiêu, nội dung và giải pháp đã đề ra tại quyết định số 667/QĐ-TTg ngày 18/5/2017 về việc phê duyệt đề án “Phát triển hệ tri thức Việt số hóa”. Bên cạnh đó, đã xác định được các loại thông tin của ngành TN&MT cần đưa lên Hệ tri thức Việt số hóa góp phần công bố, quảng bá thông tin dữ liệu ngành TN&MT đến cộng đồng, đồng thời huy động được tri thức của cộng đồng qua việc đóng góp dữ liệu mở trên Cổng dữ liệu mở của ngành TN&MT.

Các nội dung trong bài báo này được nhóm thực hiện lựa chọn, tổng hợp từ các nội dung trong đề tài NCKH cấp Bộ: “Nghiên cứu cơ sở lý luận và thực tiễn xây dựng quy định về hệ thống hóa và số hóa thông tin, tích hợp thông tin nhằm quản lý dữ liệu ngành tài nguyên và môi trường trên Hệ tri thức Việt số hóa”. Mã số: TNMT.2020.667.01. Nhóm thực hiện xin cảm ơn sự giúp đỡ, tạo điều kiện của Vụ KH&CN - Bộ TN&MT, Văn phòng đề án Hệ tri thức Việt số hóa - Bộ KH&CN đã phối hợp thực hiện, hỗ trợ kinh phí để thực hiện đề tài.

#### Tài liệu tham khảo

1. Các sản phẩm của đề tài NCKH cấp Bộ: “Nghiên cứu cơ sở lý luận và thực tiễn xây dựng quy định về hệ thống hóa và số hóa thông tin, tích hợp thông tin nhằm quản lý dữ liệu ngành TN&MT trên Hệ tri thức Việt số hóa”. Mã số: TNMT.2020.667.01;
2. Kiến trúc Chính phủ điện tử ngành TN&MT, phiên bản 2.0;
3. Kiến trúc tổng thể Hệ thống thông tin quan trắc TN&MT Quốc gia;
4. Quyết định 677/QĐ-TTg ngày 18/5/2017 của Thủ tướng Chính phủ về việc phê duyệt đề án “Phát triển Hệ tri thức Việt số hóa”;
5. Quyết định số 3166/QĐ-BTNMT ngày 12/12/2019 của Bộ trưởng Bộ TN&MT ban hành Quy chế cập nhật CSDL của Bộ TN&MT vào Hệ tri thức Việt số hóa;
6. Các tài liệu khác có liên quan. ■